AD_____

GRANT NUMBER: DAMD17-94-J-4328

TITLE: Computer-Aided Diagnosis and Automated Screening of Digital Mammogram

PRINCIPAL INVESTIGATOR: Doctor Kevin S. Woods

CONTRACTING ORGANIZATION: University of South Florida
Tampa, Florida 33620-7900

REPORT DATE: September 1996

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Annual (1 Sep 95 - 31 Aug 96)

4. TITLE AND SUBTITLE
Computer-Aided Diagnosis and Automated Screening of Digital Mammogram

5. FUNDING NUMBERS
DAMD17-94-J-4328

6. AUTHOR(S)
Doctor Kevin S. Woods

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of South Florida
Tampa, Florida 33620-7900

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

The primary objective of our on-going research is to develop computer techniques for detecting abnormalities in digital mammograms which can be used by radiologists to increase diagnostic accuracy and decrease the rate of undetected cancers. Our efforts in the second full year of research have focused on the initial stages of such a system. First, a general view of detection algorithms is presented. The basic framework involves two phases, pixel-level segmentation and region-level classification, each composed of a few elementary steps. By viewing several recent detection algorithms in the context of this general framework, we are able to show some fundamental advantages to concentrating efforts on the early pixel-level analysis of the segmentation phase. Secondly, we are developing an automated technique for quantifying and characterizing breast tissue structure. Unlike most previous efforts, the purpose of our work will be to assign a "difficulty index" to the mammogram images based on texture analysis and density of breast parenchymal tissue. The difficulty index will be incorporated as an additional parameter in a mammogram image analysis system so that decisions may be effectively conditioned on the density measure.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer , computer-aided diagnosis, digital mammography, image segmentation, density estimation.	16
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Kevin Woods 9/23/96

PI - Signature Date

Table of Contents

1 Introduction

The primary objective of our on-going research is to develop computer techniques for detecting abnormalities in digital mammograms using sound engineering methods and rigorous testing to provide reliable results which can be used by radiologists to increase diagnostic accuracy and decrease the rate of undetected cancers.

In the second full year of research, we have focused our efforts on the initial processing stages of a system for computer-aided diagnosis in digital mammography. First, after a careful examination of some of the most recent and successful techniques to mammogram image analysis, we have selected an approach that is fundamentally the same for all types of mammographic abnormalities. We provide theoretical and empirical evidence which supports our design philosophy. Second, we have begun a collaborative project with the Channing Laboratory at the Harvard Medical School and Massachusetts General Hospital to automatically estimate parenchymal tissue density, an important preprocessing step in computerized mammogram image analysis.

The following subsections will introduce the topics that have been the focus of our second full year of research. The body of this report (Section 2) will provide the details of the research conducted, including experimental methods, data, and results. Section 3 of the report will summarize the research, and draw some conclusions. Each research topic will be covered in separate subsections within each of the three major sections of this report. Much of the following material has or will be submitted to scientific journals and/or conference proceedings.

1.1 A General View of Detection Algorithms

In recent years, many techniques have been proposed for mammogram image analysis. And although they are too numerous to list here, a good historical perspective and review of computer vision and artificial intelligence in mammography can be found in a fairly recent review article [1]. Collections of work on automated mammogram image analysis include the proceedings of the First International Workshop on Digital Mammography, held in 1993 [2], and the Second International Workshop on Digital Mammography, held in 1994 [3]. Selected papers from the first conference appear in a special issue of the International Journal of Pattern Recognition and Artificial Intelligence [4]. Revisions of these papers and five additional papers appear as a collection in [5].

When examined from a certain perspective, all detection algorithms involve two basic phases: 1) segmentation of *suspicious* regions, and 2) classification of the regions as *normal* or *abnormal*. The basic segmentation and classification phases can be broken down into a few elementary steps, as in Figure 1. We would argue that any detection algorithm can be organized in this general framework.

Segmentation encompasses three elementary steps. First, one or more features are computed at every pixel. The feature(s) may be as simple as absolute pixel intensity, or more complex, such as a feature specifically designed to respond to a known image characteristic. Next, the pixels are classified as being suspicious or not. The complexity of this step may vary from a manually selected threshold on a single feature to more formal methods of statistical classification involving multiple features. Finally, suspicious pixels are organized

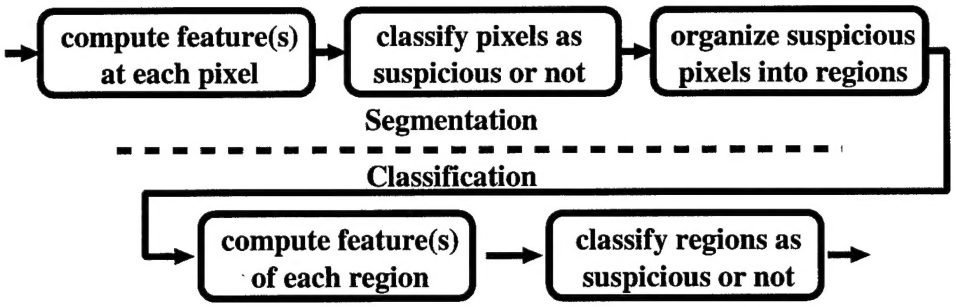

Figure 1: A general view of detection algorithms for mammogram image analysis.

into regions, usually by grouping connected pixels.

Region classification is composed of two elementary steps. First, one or more features are extracted from the suspicious regions. These features may attempt to measure properties which are not as conceptually clear at the pixel level, such as region size and shape. Finally, segmented regions are classified either *normal* or *abnormal*.

The purpose of this portion of our work is to select a fundamental design philosophy for our detection algorithms. We begin with a brief analysis of recent work, describing each algorithm in terms of the general framework of Figure 1. Next, we hypothesize why certain approaches may have fundamental advantages. These observations lead to an experiment in which we isolate the effect of improving the first step in the detection process.

1.2 Automated Characterization of Breast Tissue

An approach to a complex image analysis problem may be significantly different when some fundamental property of an image can be determined in advance. Mammogram images are 2-D projections through a highly textured 3-D structure. Mammogram interpretation involves detecting subtle changes in texture in which important detail may be obscured by tissue from above and below. Adding to the problem is the wide variation of tissue structure and background texture encountered in the images. It is well accepted that the reading of mammograms varies in difficulty according to the density of the background tissue in the image. In ours and other applications, knowing *a priori* whether an image is relatively difficult to interpret could provide advantages in subsequent processing.

We are developing an automated technique for quantifying and characterizing breast tissue structure in digital mammogram images. There have been some previous efforts for automated classification of breast parenchymal patterns and/or tissue density estimation [6, 7, 8, 9, 10]. Most of this work has been used in an attempt to show a correlation between breast density and an increased risk of breast cancer, a hypothesis offered in 1976 by Wolfe [11]. The conclusions of Wolfe are still debated today, and there have been several studies with contradictory results.

Unlike most previous efforts, the purpose of our work will be to assign a "difficulty index" to mammogram images based on texture analysis and quantification of breast parenchymal tissue. Such a difficulty index would have several potential uses in automated image analysis as well as conventional screening mammography. One, as suggested by Hajnal [10], is

for presorting mammograms into "dense" and "fatty" categories, so that the "dense" (and therefore more difficult) images can be read by a more experienced mammographer. The novel aspect of our work is the utilization of the computerized density estimation as part of an automated image analysis system. In this situation, the difficulty index can be used as an additional parameter in a classifier, so that decisions may effectively be conditioned on the density measure. Yet another potential use is as a measure of the confidence in the results of other image analysis algorithms. For example, some image analysis might be more likely to produce false positives in highly textured images.

2 Body

This section provides details of the research directed towards solving the problems introduced in the Sections 1.1 and 1.2. Whenever possible, experimental methods and results are provided.

2.1 Choosing a Fundamental Design Philosophy

By examining various techniques developed for mammogram image analysis, we may be able to determine if certain approaches have a fundamental advantage. If such an advantage can be shown via theoretical and empirical analysis, we will incorporate the appropriate techniques into the design philosophy of our mammogram image analysis software.

2.1.1 A Survey of Recent Techniques

In this section, we review five recent detection algorithms in terms of the general detection framework described in Section 1.1. These detection algorithms represent a broad range of approaches, and have been selected to illustrate a point. The algorithms and their performance are summarized in Table 1.

The University of Chicago group [12] uses a technique that finds regions of tissue that are radiographically brighter than tissue in a corresponding image through a series of adaptive grey level thresholding operations. Region-growing is employed to group suspicious pixels together, and to improve the segmentation. Thus, the pixel features used in the segmentation phase are intensity, and contrast. Here, the contrast is computed relative to pixels in a corresponding image. The classification phase involves thresholding on size, contrast, and circularity measurements of the segmented regions. On 154 mammogram image pairs (308 images) the reported sensitivity is 85% with an average of about 3.0 false positive detections per image.

Li et al. [13] use adaptive grey-level thresholding based on local contrast to get an initial segmentation, and a multiresolution Markov random field (MRF) model-based method to iteratively improve the segmentation. The mean intensity of a region surrounding a pixel guides the process. Again, the pixel features used in the segmentation phase are contrast and intensity. The classification phase uses a fuzzy binary decision tree and features based on region size, shape, contrast and smoothness. On 75 mammogram images, the reported sensitivity is 90% with an average of about 2.0 false positive detections per image.

Brzakovic et al. [14] use a multiresolution approach which is essentially an adaptive hierarchical region growing procedure. Parameter settings of the segmentation procedure determine the size and contrast of the objects that will be detected. As before, the pixel features that are utilized in the segmentation phase are contrast and intensity. A classification phase eliminates false positive detections according to their size, mean intensity, and compactness. On a set of 12 images with irregular masses, a sensitivity of about 67% is reported. On 50 normal mammogram images, an average of 0.1 false positive detections per image were reported.

Karssemeijer's [15] algorithm for detection of spiculated lesions uses two texture features which are specifically designed to detect the radiating spicule structure. In the segmentation phase, the orientation of lines are analyzed and classified using Bayesian decision theory. A classification phase reduces the number of false positives by removing detections smaller than a predetermined size. On a set of 9 images with spiculated lesions, a sensitivity of about 89% is reported. On 50 normal mammogram images, an average of 0.4 false positive detections per image were reported.

Kegelmeyer's [16] algorithm for detection of spiculated lesions uses 5 texture features and a binary decision tree classifier to label each pixel with its probability of being located on an abnormality. The result is a probability image, called a dense feature map, which is smoothed and thresholded to group pixels together for a final segmentation. There is *no* classification phase. On a set of 330 mammogram images, the reported sensitivity is 97% with an average of 0.28 false positive detections per image.

Table 1: Summary of several detection algorithms and reported performance.

Algorithm Reference	Segmentation: Pixel Features	Classification: Region Features	TP rate & FPs per image
Univ. of Chicago [12]	intensity & contrast	size, contrast, & circularity	TP rate: 85.0% FPs/image: 3.0
Li et al. [13]	intensity & contrast	shape, contrast, size & smoothness	TP rate: 90.0% FPs/image: 2.0
Brzakovic et al. [14]	intensity & contrast	size, compactness, mean intensity	TP rate: 67.0% FPs/image : 0.1
Karssemeijer [15]	2 line orientation texture measures	size	TP rate: 89.0% FPs/image : 0.4
Kegelmeyer [16]	1 line orientation & 4 general texture	none	TP rate: 97.0% FPs/image : 0.28

2.1.2 Discussion

Looking at sensitivity and average number of false positives per image, Kegelmeyer's algorithm would appear to be clearly superior, followed by Karssemeijer's algorithm. The most noticeable difference between these two algorithms and the other three is where within the detection process the majority of the "intelligence" is applied. Kegelmeyer and Karssemeijer

make extensive use of texture features and sophisticated statistical classification in the earliest possible steps of the detection algorithm. Although the other three algorithms utilize sophisticated techniques for pixel level classification (MRF model, multiresolution analysis), the early steps in the segmentation phase rely primarily on simple grey-level properties of the pixels, namely intensity and contrast measurements.

We can most likely eliminate the classification phase that each algorithm utilizes as the cause of performance differences. In fact, the two algorithms with the best reported performance have the least sophisticated methods of reducing the false positive detections which result from the segmentation phase. Karssemeijer simply eliminates small objects, while Kegelmeyer uses a null classification phase.

The above observations suggest the following. The early steps in the detection algorithm, the segmentation phase, have a much more profound effect on overall performance than steps occurring later in the classification phase. So, if the performance of the segmentation phase is poor, there may be no way to recover in the classification phase, regardless of the level of sophistication. It is not possible to improve sensitivity in the classification phase. Similarly, too many false positive regions emerging from the segmentation phase, may make it difficult for the classification phase to reduce the false positive rate to an acceptable level without having a detrimental effect on the overall sensitivity.

2.1.3 Experimental Methods and Results

Here, we present an experiment which isolates the effect of improving only the first step in the detection process. The only change from one test to another is the number and type of features extracted from each pixel. All subsequent steps remain unchanged. Brief descriptions of the data set and detection algorithm are given, although they are not of central importance. Instead, we would like to focus on the general effect that one step of a detection algorithm has on the following steps.

The dataset includes 320 images at a spatial resolution of 280 microns per pixel, 62 of which contain a visible spiculated lesion. Ground truth, which was specified by a radiologist, is denoted by a circle surrounding a lesion. The images are randomly split into two equal sets, each containing 31 abnormal images. One set is used for classifier training, and the other is used for performance evaluation.

The detection algorithm we have selected to implement is a version of Kegelmeyer's [16] dense feature map (DFM) method. The DFM approach is conceptually simple and has shown good performance in a previous application [16]. Briefly, a set of features is extracted from each pixel and organized into a feature vector. A subset of feature vectors obtained from the training images is used to grow and prune a binary decision tree. To evaluate a test image, the feature vector for every pixel is dropped into the decision tree, resulting in a "probability image" in which the pixel values represent the probability of belonging to a spiculated lesion. A spatial smoothing operation is performed to achieve a consensus among neighboring pixels. A final segmentation is produced by thresholding the probability image. More detail on the DFM method can be found in [16].

The performance metrics we report are: 1) the true positive rate and average number of false positives per image, 2) the mean false positive area segmented per image (i.e. amount of normal tissue mislabeled), and 3) the average ratio of the area of a true positive detection to

the area of the ground truth circle. This third metric helps determine how well abnormalities are segmented, in general. A detection is considered a true positive if at least 50% of the detection is part of the ground truth circle. This prevents large detections which happen to overlap a ground truth circle from being labeled true positives.

Table 2: Description of pixel features extracted in the segmentation phase.

Feature	Description
Pixel Intensity	Absolute pixel intensity averaged over a $5mm$ window
St. dev of Intensity	Standard deviation of pixel intensity in a $5mm$ window
Size estimate	Local contrast for several window sizes is used to estimate the size of the region a pixel *may* belong to.
ALOE	Analysis of local edge orientation [16] over a $2cm$ window
Extrema Density	Texture measure of "roughness" over a $1cm$ window
Local contrast of texture	Local contrast of a smoothed edge gradient image estimated from several window sizes
Average gradient	Edge gradient image (Sobel) smoothed with a $5mm$ window
Average gradient	Edge gradient image (Sobel) smoothed with a $5mm$ window (raw image is preprocessed to remove low frequency background)
St. dev of Intensity	Standard deviation of pixel intensity in a $5mm$ window (raw image is preprocessed to remove low frequency background)

Beginning with three intensity features, the detection algorithm is trained and evaluated. The algorithm is retrained and evaluated several times, each time adding another texture feature to the first step of the segmentation phase. The initial set of three features is meant to roughly correspond to the intensity and contrast features used as the basis for segmentation in the first three algorithms listed in Table 1. The features are listed in Table 2 in the order they were added to the detection algorithm.

Experimental results are summarized in Table 3. For each instance of the detection algorithm, the probability images are thresholded such that the maximum sensitivity is achieved. For example, using only three intensity features as the basis for segmentation, only 48% of the lesions could be segmented. Raising the threshold would result in a lower sensitivity as fewer pixels survive the thresholding. Lowering the threshold would also lower the sensitivity as more pixels survive the threshold step, but the segmented regions become too large and do not pass as a true positive according to our performance metric. A general trend emerges. As texture features are added, the algorithm's sensitivity improves while the average amount of false positive tissue segmented decreases. As expected, the algorithm's performance eventually begins to decline as more features are added.

2.2 High-level Image Characterization Preprocessing

The percent of the breast tissue that appears dense on a mammogram reflects the proportion of stromal and epithelial tissue compared to fat, and varies considerably among healthy

Table 3: Segmentation results for one algorithm with various features used as input.

Pixel Features Used for Segmentation	Percent of Normal Tissue Misclassified	Average Ratio of TP Area to Ground Truth	TP fraction & FPs per Image
3 intensity	2.8%	0.50	48% TPF 2.0 per image
3 intensity, 1 texture	1.9%	0.53	61% TPF 1.4 per image
3 intensity, 2 texture	1.8%	0.58	74% TPF 1.8 per image
3 intensity, 3 texture	**1.9%**	**0.57**	**84% TPF 1.8 per image**
3 intensity, 4 texture	2.0%	0.60	71% TPF 2.2 per image
3 intensity, 6 texture	2.0%	0.46	64% TPF 2.4 per image

women. Most women have some mammographic density, with the majority of women having more than 25% of their breast comprised of dense tissue. Increased mammographic density decreases the sensitivity and specificity of mammographic detection of breast cancer, and there is no reason to believe this would not be the case with computerized interpretation of these images. Simply put, a mammographic abnormality embedded in dense breast tissue is more difficult to detect than one surrounded by fatty tissue. Abnormalities embedded in dense connective tissue are more radiographically subtle, and will likely respond to image processing operators differently than nearly identical abnormalities surrounded by more radiolucent fatty tissue.

As described in the sections of this report devoted to the fundamental approach of detection algorithms, the first step of such an algorithm is to compute useful features for each pixel in the image. For reasons just described, it will be important to know if a pixel, which represents a small area of breast tissue, is embedded in dense or fatty tissue. We can envision at least two possible uses for this type of information. First, fatty and dense breasts can be pre-sorted such that each type is processed by separate detection algorithms which have been fine-tuned to take into account the composition of the breast. Second, the density associated with a pixel is a distinct feature which can be input to the detection algorithm and processed as any other feature. In this situation, tissue density is incorporated into the statistical models used to describe normal and abnormal breast tissue. It is possible that a combination of the two approaches will be useful.

Preliminary results of our density estimation technique have been shown to radiologists and others involved in breast cancer research, and their responses have been encouraging. Figure 2 shows an example of these results. As we are in the early stages of this work, there are no final test results at the present. The following subsections outline the experimental procedures we are following for the development and testing of our approach.

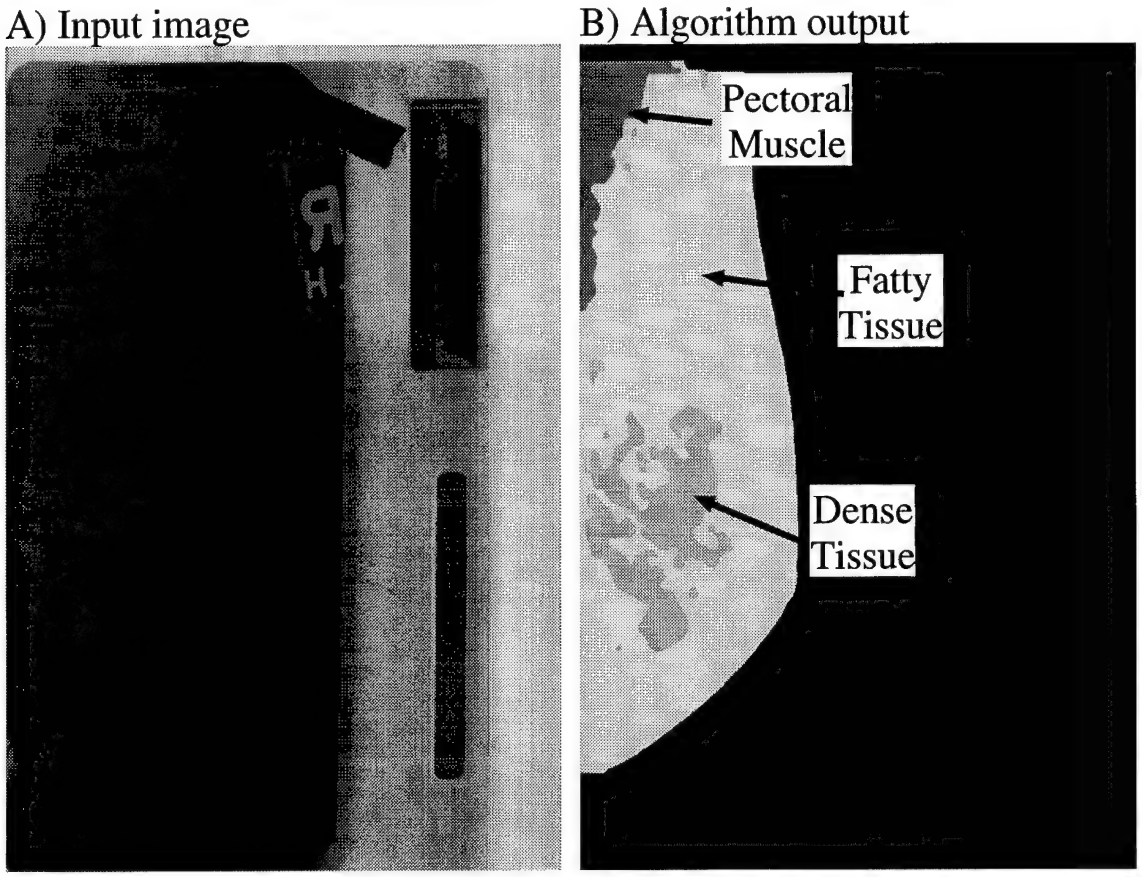

Figure 2: Preliminary results of a technique for automated breast segmentation and dense tissue estimation. A) The mammogram image used as input. This image has been enhanced for display purposes. B) The output of the algorithm. The breast region has been segmented from the background, and the dense tissue, fatty tissue and pectoral muscle have been segmented and labeled.

2.2.1 Experimental Data and Ground Truth

In order to be of practical value, we require the breast density assessment of our automated technique to closely match that of an experienced breast radiologist. Thus, we begin with a set of mammograms from twenty 4-view cases, a total of 80 images, which were manually selected and digitized. The cases were selected to represent the wide range of breast tissue density that would occur in a typical screening program. Next, a radiologist marked the regions of dense tissue on each mammogram with a grease pencil, and the films were re-digitized. The films were cleaned, marked by a second radiologist, and digitized a third time. The marked images were used to create ground truth templates by manually tracing the radiologists' markings with a computer mouse. The ground truth templates are overlays of the raw mammogram image which denote regions corresponding to dense tissue.

So, we have three images of each mammogram: 1) the raw image to be processed by our automated technique, 2) ground truth denoting dense tissue regions as estimated by a radiologist, and 3) another ground truth image estimated by a second radiologist. The

computer output is a template image denoting the regions of dense tissue as estimated by our automated technique. Distinct "landmarks" were placed on the original films prior to digitization such that the three template images (2 ground truth and 1 computer output) of the same film can be registered and aligned. This permits accurate comparisons of the density estimations made by the three "experts". For example, after aligning the templates we can compute the intersection and the union of the dense tissue estimates from the two sources. Dividing the intersection of areas by the union of areas results in a "measure of agreement" which ranges from 0, for no agreement, to 1, for a perfect match.

The purpose of having ground truth estimated by two radiologists is so that a tolerance for the computerized scheme can be determined. A measure of the inter-observer variability between the two radiologists over the set of mammograms is computed. This same measure of variability will be computed between the automated technique and one (or possibly both) of the radiologists. We can consider the performance of the automated technique to be satisfactory when the variability measures are nearly equal. Thus, we will require the density estimation of our automated technique to match that of an experienced breast radiologist as closely as we would expect the estimate of another experienced breast radiologist to match.

2.2.2 Experimental Methods

Our basic approach to image segmentation is Kegelmeyer's dense feature map, described in Section 2.1.3. The features extracted from each pixel, and the method of pixel classification need to be determined. A large pool of texture features, and several methods of statistical classification will be examined. The goal is to find the fewest number of features and the most simple method of classification that will achieve a satisfactory level of performance.

More specifically, the computer algorithm will read in a raw mammogram image, segment the breast tissue from the film background (extracting the pectoral muscle if necessary), and classify the breast tissue as either fatty or dense. Thus, the breast region in the image is segmented from the background in the same step that tissue density is estimated. This means we will require ground truth for classifier training from four types of image regions: 1) dense breast tissue, 2) fatty breast tissue, 3) pectoral muscle, and 4) film background.

Since the radiologists have only provided ground truth for the dense tissue regions of the mammogram image, we added ground truth for three other region types. Fatty tissue is simply the remaining breast tissue that was not been labeled as dense tissue by either radiologist. The film background is anything not considered breast tissue (fatty or dense) or pectoral muscle. We should note that our ground truth for fatty tissue, pectoral muscle, and film background does not require the same precision or radiological expertise as we require for the dense tissue ground truth. This is true since the ground truth for these three other regions is only used for classifier training. Only ground truth for the dense tissue regions is used to assess algorithm performance.

In order to select the classifier and features used in the automated density estimation algorithm, the set of 20 cases are divided into two equal halves. One half of the data is used as training data to learn system parameters, and the other half is used as an independent test set to evaluate system performance. Next, the roles of the two sets are reversed. In this way, we get unbiased test results for all images in the data set. Using the tolerance defined by comparing the performance of the two radiologists, the feature set and classifier to be

used in the final version of our automated system are selected. All other things being equal, the final system configuration selected is the one with the fastest execution time.

3 Conclusions

This section summarizes our results and analysis. We indicate the implications this work has on our future efforts and in obtaining the goals set forth in the original proposal.

3.1 An Approach to Mammogram Image Analysis

A general view of detection algorithms is presented. The basic framework involves two phases, pixel level segmentation and region level classification, each composed of a few elementary steps. By viewing several detection algorithms in the context of this general framework, we are able show some fundamental advantages to concentrating efforts on the early pixel level analysis of the segmentation phase.

First, the performance of one step in a detection algorithm is dependent on the performance of the previous step. For example, the features extracted from each pixel affect the results of the pixel level classification, and therefore, the overall segmentation. Since there is a cumulative effect in which one step in the detection algorithm is directly affected by the performance of the previous step, our future research will concentrate on improving our current detection algorithms by concentrating on the earliest steps in the process.

Another fundamental advantage of focusing on the early steps of pixel level analysis is the amount of data available for classifier training. Extracting features from each pixel provides hundreds of thousands of training samples to characterize *normal* and *abnormal* tissue. After pixels have been grouped together, the number of regions available for classifier training is usually a hundred or so at best. Since there are orders of magnitude more samples available at the pixel level, more statistically accurate and robust measures of image features can be obtained in the earliest phases of a detection algorithm.

3.2 Breast Density Estimation

We are in the process of developing an automated technique for quantifying and characterizing breast tissue structure in digital mammogram images. Once the density estimation technique is performing as well as a trained radiologist, the plan is to incorporate the density estimation, possibly as a difficulty index, into our mammogram image analysis algorithms and evaluate the system performance. The density estimation will be used in two ways: first, as an additional parameter for the classification problem, and second, as a method for sorting images prior to classifier training and testing. Eventually, we plan to use classification results and the difficulty index to determine if there is a correlation between classification accuracy and the perceived difficulty of an image. The objective here is to determine if the difficulty index can be incorporated into a confidence value associated with the system output.

References

[1] C. J. Vyborny and M. L. Giger, "Computer vision and artificial intelligence in mammography," *AJR*, vol. 162, pp. 699–708, 1994.

[2] SPIE, *Biomedical Image Processing and Biomedical Visualization*, vol. 1905, (San Jose, CA), SPIE - The International Society for Optical Engineering, February 1-4 1993.

[3] A. G. Gale, S. M. Astley, D. R. Dance, and A. Y. Cairns, eds., *Digital Mammography: Proceedings of the 2nd International Workshop on Digital Mammography*, vol. 1069 of *International Congress Series*, (York, England), Elsevier Science B. V., July 10-12 1994.

[4] K. W. Bowyer and S. Astley, eds., *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7(6). World Scientific Publishing Company, December 1993. Special Issue: State of the Art in Digital Mammographic Image Analysis.

[5] K. W. Bowyer and S. Astley, eds., *State of the Art in Digital Mammographic Image Analysis*. Singapore: World Scientific Publishing Company, 1994.

[6] C. Kimme-Smith, G. Frankl, and G. W. adn J Slansky, "Toward reliable measurements of breast parenchymal patterns," *IEEE Computer Applications in Radiology and Analysis of Radiological Images*, vol. 1979, pp. 118–121, VI.

[7] I. E. Magnin, F. Cluzeau, and C. L. Odet, "Mammographic texture analysis: An evaluation of risk for developing breast cancer," *Optical Engineering*, vol. 25, no. 6, pp. 780–784, 1986.

[8] C. B. Caldwell, S. J. Stapleton, D. W. Holdsworth, R. Jong, W. Weiser, G. Cooke, and M. J. Yaffe, "Characterization of mammographic parenchymal patterns by fractal dimension," *SPIE Vol. 1092 Medical Imaging III: Image Processing*, vol. 1092, pp. 10–16, 1989.

[9] P. Miller and S. Astley, "Classification of breast tissue by texture analysis," *Image and Vision Computing*, vol. 10, no. 5, pp. 277–282, 1992.

[10] S. Hajnal, P. Taylor, M. Dilhuydy, and B. Barreau, "Classifying mammograms by density: sorting for screening," in *State of the Art in Digital Mammographic Image Analysis* (K. W. Bowyer and S. Astley, eds.), pp. 64–81, Singapore: World Scientific Publishing Company, 1994.

[11] J. N. Wolfe, "Breast patterns as an index of risk for developing breast cancer," *American Journal of Roentgenology*, vol. 126, pp. 1130–1139, 1976.

[12] R. M. Nishikawa, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms," in *State of the Art in Digital Mammographic Image Analysis* (K. Bowyer and S. Astley, eds.), pp. 82–102, Singapore: World Scientific Publishing Company, 1994.

[13] H. Li, M. Kallergi, L. Clarke, V. Jain, and R. Clark, "Markov random field for tumor detection in digital mammography," *IEEE Trans. on Medical Imaging*, vol. 14, pp. 565–576, September 1995.

[14] D. Brzakovic and M. Neskovic, "Mammogram screening using multiresolution-based image segmentation," in *State of the Art in Digital Mammographic Image Analysis* (K. Bowyer and S. Astley, eds.), pp. 103–127, Singapore: World Scientific Publishing Company, 1994.

[15] N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clusters in mammograms," in *State of the Art in Digital Mammographic Image Analysis* (K. Bowyer and S. Astley, eds.), pp. 148–166, Singapore: World Scientific Publishing Company, 1994.

[16] W. Kegelmeyer, J. Pruneda, P. Bourland, A. Hillis, M. Riggs, and M. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiology*, vol. 191, pp. 331–337, 1994.